DOCUMENT RESUME

ED 414 939 IR 056 797

AUTHOR Neff, Raymond K.
TITLE A New Consortial Model for Building Digital Libraries.
PUB DATE 1997-04-00
NOTE 21p.; Paper presented at the Conference on Scholarly Communication and Technology (Atlanta, GA, April 24-25, 1997), see IR 056 774.
AVAILABLE FROM Association of Research Libraries (ARL) Web site: http://www.arl.org/scomm/scat/
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Libraries; Consortia; Electronic Libraries; Higher Education; Information Networks; *Information Systems; *Library Networks; *Research Libraries; Resource Allocation; *Scholarly Journals; *Shared Resources and Services
IDENTIFIERS *Case Western Reserve University OH; Digital Data

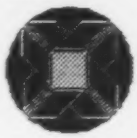
ABSTRACT

The libraries in U.S. research universities are being systematically depopulated of current subscriptions to scholarly journals. Annual increases in subscription costs are consistently outpacing the growth in library budgets; this has become a chronic problem for academic libraries which collect in the fields of science, engineering, and medicine. Case Western Reserve University has built a novel digital library distribution system and focused collections in the chemical sciences to investigate a new approach to solving a significant portion of this problem. By collaborating with another research library which has a strong chemical sciences collection, they have developed a methodology to control costs of scholarly journals and have planted the seeds of a new consortial model for building digital libraries. This paper summarizes Case Western Reserve University's progress to date and indicates areas in which the University is continuing its research and development. The University's rights management system, consortial and equipment standards, scanning and workflow, and technical justification for a digitization for the consortium are described in appendices. (Author/AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 414 939

Scholarly Communication and Technology



Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #7 Multi-Institutional Cooperation

A New Consortial Model for Building Digital Libraries

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Richard Ekman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Raymond K. Neff
Vice President for Information Services
Director of University Libraries
Case Western Reserve University

rkn@po.cwru.edu

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

The libraries in America's research universities are being systematically depopulated of current subscriptions to scholarly journals. Annual increases in subscription costs are consistently outpacing the growth in library budgets. This has become a chronic problem for academic libraries which collect in the fields of science, engineering, and medicine, and by now the problem is well recognized (Cummings, 1992). At Case Western Reserve University, we have built a novel digital library distribution system and focused on our collections in the chemical sciences to investigate a new approach to solving a significant portion of this problem. By collaborating with another research library which has a strong chemical sciences collection, we have developed a methodology to control costs of scholarly journals and have planted the seeds of a new consortial model for building digital libraries. This paper summaries our progress to

IR056797

BEST COPY AVAILABLE

date and indicates areas in which we are continuing our research and development.

For research libraries in academia, providing sufficient scholarly information resources in the chemical sciences represents a large budgetary item. For our purposes, the task of providing high-quality library services to scholars in the chemical sciences is similar to providing services in other sciences, engineering, and medicine; if we solve the problem in the limited domain of the chemical sciences, one can reasonably extrapolate our results to these other fields. Thus, research libraries whose mission it is to provide a high level of coverage for scholarly publications in the chemical sciences are the focus of this project, although we believe that the principles and practices employed in this project are extensible to the serial collections of other disciplines.

A consortium depends on having its members operating with common missions, visions, strategies, and implementations. We adopted the tactics of developing a consortial model by having two neighboring libraries collaborate in the initial project. The University of Akron (UA) and Case Western Reserve University (CWRU) both have academic programs in the chemical sciences which are nationally ranked, and the two universities are fewer than thirty miles apart. It was no surprise to find that both universities have library collections in the chemical sciences which are of high quality and nearly exhaustive in their coverage of scholarly journals. To quantify the correlation between these two collections we counted the number of journals which both collected and found the common set to be 76% in number and 92% in cost. The implications of the overlap in collecting patterns is plain; if both libraries collected only one copy of each journal, with the exception of the most used journals, approximately half of the cost of these subscriptions could be saved. For these two libraries, the cost savings is potentially $400,000 per year. This seemed like a goal worth pursuing, but to do so would require building a new type of information distribution system.

The reason scholarly libraries collect duplicative journals is that students and faculty want to be able to use these materials by going to the library and looking up a particular volume or by browsing the current issues of journals in their field. Eliminating a complete set of the journals at all but one of our consortial libraries would deprive local users of this walk-up-and-read service. We asked ourselves if it would be possible to construct a virtual version of the paper-based journal collection which would be simultaneously present at each consortium member institution, allowing any scholar to consult the collection at will even though only one copy of the paper journal was on the shelf. The approach we adopted was to build a digital delivery system that would provide to a scholar on the campus of a consortial member institution, on a demand basis, either a soft or hard copy of any article for which a subscription to the journal was held by a consortial member library. Thus, according to this vision, the use of information technology would make it possible to collect one set of journals among the consortium members and to have them simultaneously available at all institutions. Although the cost of building the new digital distribution system is substantial, it was considered as an experiment worth undertaking. The generous support of The Andrew W. Mellon Foundation is being used to cover approximately one-half of the costs for the construction and operation of the digital distribution system, with Case Western Reserve University covering the remainder. The University of Akron Library has contributed its expertise and use of its chemical sciences collections to the project.

It also seemed necessary to us to want to invite the cooperation of journal publishers in a project of this kind. To make a digital delivery system practical would require having the rights to store the intellectual property in a computer system, and when we started this project, no

consortium member had such rights. Further, it was both the on-going publications and the "back files" which would be needed so that complete "runs" of each serial could be constructed in digital form. The publishers could work out agreements with the consortium to provide their scholarly publications for inclusion in a digital storage system which would be connected to our network-based transmission system, and thus, their cooperation would become essential. The chemical sciences are disciplines in which previous work with electronic libraries had been started. The TULIP Project of Elsevier Science (TULIP, 1996) and the CORE Project of Cornell University, the American Chemical Society, Bellcore, Chemical Abstracts, and OCLC were known to us, and we certainly wanted to benefit from their experiences. Publications of Elsevier Science, the American Chemical Society, and others including Springer-Verlag, the Academic Press, and John Wiley & Sons were central to our proposed project because of the importance of their journal titles to the chemical sciences disciplines.

We understood from the beginning of this effort that we would want to monitor the performance of the digital delivery system under realistic usage scenarios. The implementation of our delivery system has built into it extensive data collection facilities for monitoring what users actually do. The system is also sensitive to concerns of privacy in that it collects no items of performance information which may be used to identify unambiguously any particular user.

Given the existence of extensive campus networks at both CWRU and UA and substantial internetworking among the academic institutions in northeastern Ohio, there was sufficient infrastructure already in place to allow the construction and operation of an intra- and intercampus digital delivery system. Such a digital delivery has now been built and made operational. The essential aspects of the digital delivery system will now be described.

A Digital Delivery System

The roots of the electronic library are found in landmark papers by Bush (1945) and Kemeny (1962). Most interestingly, Kemeny foreshadowed what the prospective scholarly users of our digital library told us as their requirement that they be able to see each page of a scholarly article preserved in its graphical integrity. That is, the electronic image of each page layout needed to look like it did when originally published on paper. The system we have developed uses the ACROBAT^{R} page description language to accomplish this objective.

Because finding aids and indices for specialized publications are too limiting, users also have the requirement that the article's text be searchable with limited or unlimited discipline-specific thesauri. Our system complements the page images with an optical character-recognition (OCR) scanning of the complete text of each article. In this way, the user may enter words and phrases the presence of which in an article would constitute a "hit" for the scholar.

One of the most critical design goals for our project was the development of a scanning subsystem that would be easily reproducible and cost efficient to set up and operate in each consortium member. Not only did the equipment need to be readily available, but it had to be adaptable to a variety of work-flow and staff work patterns in many different libraries. Our initial design has been successfully tailored to the needs of both the CWRU libraries and the Library at the University of Akron. Our approach to the sharing of paper-based collections is to use a scanning device to copy the page images of the original into a digital format which may be readily transmitted across our existing telecommunications infrastructure. In addition, the digital version of the paper original may be stored for subsequent retrieval. Thus, repeated viewing of

the same work would necessitate only a one-time transformation of format. This is both an advantage in achieving faster response times for scholars but promotes the development and use of quality control methods. The scanning equipment we have used in this project is the Minolta PS-3000 Digital Planetary Scanner with the Epic 3000 Software Subsystem. The principal advantage of this scanner is that bound serials may be scanned without damaging the volume and without compromising the resulting page images; in fact, the original journal collection remains intact and accessible to scholars throughout the project. This device is also sufficiently fast that a trained operator, including students, may scan over 800 pages per average workday. For a student worker making $7.00 per hour, the person-cost of scanning is under $0.07 per page; the cost of conversion to searchable text adds $0.01 per page. Thus, each consortium member would be expected to make a reasonable investment in equipment, training, and personnel. Appendix D gives more details regarding the scanning processes and workflow. Appendix E gives a technical justification for a digitization standard for the consortium.

The target equipment for viewing an electronic journal was taken to be a common PC-compatible computer workstation, hereafter referred to as a client. This client is also the user platform for the on-line library catalog systems found on our campuses, as well as the growing collections of CD-ROM-based information products. Appendix C gives the specification of the workstation standards for this project. The implications for use of readily available equipment is that the client platform for our project would also work outside of the library - in fact, wherever a user wanted to work. Therefore, by selecting the platform we did, we extended the project to encompass a full campus-wide delivery system. Because our consortium involves multiple campuses (two at the outset), the delivery system is general purpose in its availability as an access facility.

Just as we had within the classical research library a place to store paper-based journals, we needed to specify a place to storage the digital copies. In technical parlance, this storage facility is called a server. To give us the greatest possible flexibility in developing the project, we decided to form the server out of two interlinked computer systems, a standard IBM System 390 with the OS/390 Open Edition version as the operating system and a standard IBM RS/6000 System with the AIX version of the UNIX operating system. Both of these components may be incrementally grown as the project's server requirements increase. Both systems are relatively commonplace at academic sites, although only one systefn pair is needed in this project, and to provide for both reliability and load leveling, it is likely that eventually two pairs of systems would be needed for an effort on the national scale.

The campus-wide networks on both our campuses and the state-wide network which connects to them uses the standards-based TCP/IP protocols. Thus, any connected client workstation which follows our minimum standards will be able to use the digital delivery system being constructed. Because the key to minimizing the operating costs within a consortium is interoperability and standardization of equipment, we have adopted a series of standards for this project; they are given in Appendices B and C. The minimum transmission speed on the CWRU campus is ten million bits-per-second (M bps) to each client workstation and a minimum of 155 M bps on each backbone link. The principal document repository is on the IBM System 390 which uses a 155 M bps ATM (asynchronous transfer mode) connection to the campus backbone. The linkage to the University of Akron is by way of the state-wide network where the principal backbone connection from CWRU is also operating at 155 M bps, and the linkage from the UA to the state-wide network is at 3 M bps. The on-campus linkage for UA is also a minimum of 10 M bps to each client workstation within the chemical sciences scholarly community and to client workstations in the UA University Library.

One of the most significant problems in placing intellectual property in a networked environment is that with a few clicks of a mouse thousands of copies of the original work can be distributed at virtually zero marginal cost, and the owner is generally deprived of expected royalty revenue. Since we recognized this problem some years ago and we realized that solutions outside of the network itself were unlikely to be either permanent or satisfactory to all parties (e.g., author, owner, publisher, distributor, user), we embarked on the creation of a software subsystem now known as Rights Manager™. With our RM system, we can control the dissemination of network-based intellectual property subject to each stakeholder receiving his due. Appendix A gives a fuller description of the RM system.

The key to understanding our approach to intellectual property management is that we expect that each scholarly work will be disseminated according to a comprehensive contractual agreement. Publishers may use master agreements to cover a set of titles. Further, we do not expect that there will be only one interpretation of concepts such as "fair use," and our Right Manager system makes provision for arbitrarily different operational definitions of fair use, so that specific contractual agreements can be "enforced" within the delivery system.

A New Consortial Model

The library world has productively used various consortial models for over thirty years, but until now, there has not been a successful model for building a digital library. One of the missing pieces in the consortial jigsaw puzzle has been a technical model which is both comprehensive and reproducible in a variety of library contexts. To begin our approach to a new consortial model, we developed a complete technical system for building and operating a digital library. Building such a system is no small achievement. Similar efforts have been undertaken with the Elsevier Science TULIP Project and the JSTOR project.

The primary desiderata for a new consortial model are as follows:

- Any research library can participate using agreed upon and accepted standards.
- Many research libraries each contribute relatively small amounts of labor by scanning a small, controlled number of journal issues. Scanning is both systematic and based on a request for an individual article.
- Use of readily available off-the-shelf equipment.
- Intellectual property is made available through licensing and controlled by the Rights Manager software system.
- Publishers grant rights to libraries to scan and store intellectual property retrospectively (i.e., already purchased materials) in exchange for the right to license use of the digital formats to other users. Libraries provide publishers with digital copies of scholarly journals for their own use, thus enabling publishers to enrich their own electronic libraries.

A Payments System for the Consortium

It is unrealistic to assume that all use of a future digital library will be without any charging mechanisms even though the research library of today charges for little except for photocopying and user fines. This is not to assume that the library user is charged for each use although that would be possible. More likely it would be the library which would pay on behalf of the members of the scholarly community (i.e., student, professor, researcher) it supports. According to our proposed consortial model, libraries would be charged for use of the digital library according to the total pages "read" in any given user session. It could be easily worked out such that users who consult the digital library on the premises of the campus library would not be charged themselves, but if they used the digital library from another campus location or from off-campus through a network, that they would pay a per-page charge analogous to the cost of photocopying. A system of charging could include categorization by type of user, and the RM system provides for a wide variety of charging models, including the making of distinctions of usage in soft copy format, hard copy format, and downloading of a work in whole or in part. Protecting the rights of the owner is an especially interesting problem when the entire work is downloaded in a digital format. Both visible and invisible watermarking are techniques with which we have experience for protecting rights in the case of downloading an entire work.

We also have in mind that libraries which provide input via scanning to the decentralized, digital library would receive a credit for each page scanned. It is clear that the value of the digital library to the end user will increase as higher degrees of completeness in digitized holdings is achieved. Therefore, the credit system to originating libraries should recognize this and reward these libraries according to a formula that charges and credits with a relative credit-to-charging ratio of perhaps in the neighborhood ten to one; that is, an originating library might receive a credit for scanning equal to a charge for ten soft copy reads.

The charge-and-credit system for our new consortial model is analogous to that used for the highly successful Online Computer Library Center's cataloging system. Member libraries within OCLC contribute original cataloging entries in the form of MARC records for the OCLC database as well as draw down a copy of a holding's data to fill in entries for their own catalog systems. The system of charging for "downloads" and crediting for "uploads" is repeated in our consortial model for retrospective full-text journal articles. Just as original cataloging is at the heart of OCLC, original scanning is at the heart of our new consortial model for building the library of the future.

Data Collection

One of the most important aspects of this project is that we have instrumented the entire software system which underlies the project with data collection points. In this way we can find out through actual usage by faculty, students, and research staff what aspects of the system are good and which need more work and thought. Over the past decade many people have speculated about how the digital library might be made to work for the betterment of scholarly communications. The system described in this paper is one of the most comprehensive attempts yet to have experience benefit visioning.

To appreciate the detailed data being collected by the project, we will describe the various types of data that the RM system captures. Many types of transactions occur between the RM client and the server software throughout a user session. The server software record these transactions to permit detailed analysis of usage patterns. A typical user session generates the following transactions between client and server.

- User requests an article (usually from a Web browser)

  If the user is starting a new session, the RM system downloads and launches the appropriate viewer which will process only encrypted transactions. In the case of Adobe Acrobat, the system downloads a plug-in. The following transactions take place with the server:

  1a. Authenticate the viewer (i.e., ensure we are using a secure viewer).

  1b. Get permissions (i.e., obtain a set of user permissions, if any. If it is a new session, the user is set by default to be the general-purpose category of PUBLIC).

  1c. Get Article (download the requested article. If step b returns no permissions, this transaction does not occur. The user must sign on and request the article again).

- User signs on

  If the general user has no permissions, s/he must log on. Following a successful logon, transactions 1b and 1c must be repeated. Transactions during sign-on include:

  2a. Sign On

- Article is displayed on screen

  Before an article is displayed on the screen, the viewer enters an step-by-step RM process or protocol wherein a single reporting command is sent to the server several times with different state flags and use types. RM events are processed similarly for all supported functions, including display, print, excerpt, and download. The transactions include:

  3a. Report Use BEGIN (Just before the article is displayed).

  3b. Report Use ABORT (Sent in the event that a technical problem prevents display of the article (such as out of memory, etc.)).

  3c. Report Use DECLINE (Sent if the user declines display of the article after seeing the cost).

  3d. Report Use COMMIT (Just after the article is displayed).

  3e. Report Use END (Sent when the user dismisses the article from the screen by closing the article window).

- Users closes viewer

  When a user closes a viewer, an end-of-session process occurs which sends transaction (3e) for all open articles. Also a close viewer transaction is sent which immediately expires the viewer so it may not be used again.

  4a. Close Viewer

The basic data being collected for every command (with the exception of 1a) and being sent to the server for later analysis includes the following:

- Date/Time
- Viewer ID
- User ID (even if it is PUBLIC)
- IP Address of request

These primary data may be used to derive additional data: Transaction (1b) may be effectively used to log unsuccessful access attempts, including failure reasons. The time interval between transactions (3a) and (3e) may be used to measure the duration that an article is on the screen. The basic data collection module in the RM system is quite general and may be used to collect other information and derive other measures of system usage.

Conclusions

A digital distribution system for storing and accessing scholarly communications has been constructed and installed on the campuses of Case Western Reserve University and the University of Akron. This low-cost system can be extended to other institutions with similar requirements because the system components, together with the way they have been integrated, were chosen to facilitate the diffusion of these technologies. This distribution system successfully separates ownership of library materials from access to them.

The most interesting aspect of the new digital distribution system is that it can be the basis for libraries to form consortia which can share highly specialized materials, rather than duplicating them in parallel, redundant collections. When a consortium can share a single subscription to a highly specialized journal, then we have the basis for reducing the total cost of library materials because we can eliminate duplicative subscriptions. We believe that the future of academic libraries points to the maintenance of a basic core collection, the selective acquisition of specialty materials, and the sharing across telecommunications networks of standard scholarly works. The consortial model which we have built and tested is one way to accomplish this goal.

Our approach is contrasted with the common behavior of building up ever larger collections of standard works, so that over time, academic libraries begin to look ever more alike in their collecting habits and offer almost duplicative services and require ever larger budgets. This project is attempting to find another path.

The effects of the new consortial model for building digital libraries are not confined to the domain of technology. During the period when the new digital distribution system was being constructed, an agency of the Ohio Board of Regents called OhioLINK commenced an overlapping experiment with Elsevier Science. According to this recently signed agreement, all of Elsevier Science's eleven-hundred-plus electronic journals will be available for access and use on all of the 55 campuses of OhioLINK member institutions, including CWRU and the University of Akron. The cost of the entire collection of electronic journals for each university for 1997 was set by the OhioLINK contract to be approximately 5.5% greater than the institution's Elsevier Science expenditure level for 1996 subscriptions regardless of the particular subset these subscriptions represented; there is a further 5.5% price increase set to take effect in 1998. Further, the agreement between OhioLINK and Elsevier constrains the member

institutions to pay for this comprehensive access even if they cancel a journal subscription. Notably, there is an optional payment discount of 10% when an existing journal subscription (in a paper format) is limited to electronic delivery only (eliminating the delivery of a paper version). Thus, electronic versions of the Elsevier journals which are part of our chemical sciences digital library will be available at both institutions regardless of the existence of our consortium; pooling collections according to our consortial model would be a useless exercise from a financial point of view.

Other publishers are also working with our consortium of institutions to offer digital products. During spring 1997, CWRU and the University of Akron entered into an agreement with Springer-Verlag to evaluate their offering of fifty or so electronic journals, some of which overlapped with our chemical sciences collection. In 1996, OhioLINK also worked out an agreement on behalf of its member institutions with Academic Press to offer their collection of approximately 175 electronic journals, many of which were in our chemical sciences collections. Significantly, the OhioLINK contract with Academic Press facilitated the development of our digital library because it included a provision covering the scanning and storage of retrospective collections (i.e., "backfiles") of their journals which we had originally acquired by subscription. A similar agreement covering backfiles of Elsevier journals is currently under negotiation. During the development of this project, we had numerous contacts with the American Chemical Society with the objective of including their publications in our digital library. Indeed, the outline of an agreement with them was discussed. As the time came to render the agreement in writing, they withdrew and later disavowed any interest in a contract with the consortium. At the present time, discussions are being held with other significant chemical science publishers about being included in our consortial library. This is clearly a dynamic period in journal publishing and each of the societal and commercial publishers sees much at stake. While we in universities try to make sense of both technology and information service to our scholarly communities, the publishers are each trying to chart their own course both competitively and strategically while improvements in information technology continually raise the "ante" for continuing to stay in the "game."

Over the past decade several interesting experiments have been conducted to test different ideas for developing digital libraries, and more are under way. With many differing ideas and visions, an empirical approach is a sound way to make progress from this point forward. Our consortium model with its many explicit standards and integrated technology seems to us to be an experiment worth continuing. During the next few years it will surely develop a base of performance data which should provide insights for the future. In this way, experience will benefit visioning.

REFERENCES:

Borghuis, M., Brinckman, H., Fischer, A., Hunter, K., van der Loo, E., Mors, R., Mostert, P., and Zilstra, J.: TULIP Final Report: *T*he *U*niversity *Li*censing *P*rogram. New York: Elsevier Science, 1996.

Bush, V.: "As We May Think" The Atlantic Monthly, 176, 101-108, 1945.

Cummings, A.M., Witte, M.L., Bowen, W.G., Lazarus, L.O., Ekman, R.H.: University Libraries and Scholarly Communication: A Study Prepared for The Andrew W. Mellon Foundation. The Association of Research Libraries, 1992.

Fleischhauer, C. and Erway, R.L.: "Reproduction-Quality Issues in a Digital-Library System: Observations on the Reproduction of Various Library and Archival Material Formats for Access and Preservation." An American Memory White Paper, Washington, D.C.: Library of Congress, 1992.

Kemeny, J.G.: "A Library for 2000 A.D." in Greenberger, M. (Ed.), Computers and the World of the Future. Cambridge, MA.: The M.I.T. Press, 1962.

Appendix A: Rights Manager™

Case Western Reserve University has developed a rights management system (called Rights Manager™) for controlling the distribution of digitally formatted intellectual property in a networked environment. This appendix is a high-level description of the system.

CWRU has been working for the past seven years to address various problems in building a digital library. During this period, it has collaborated on a variety of projects involving multimedia authoring and presentation software systems; however, its primary objective has been the development of a client server-based content delivery system that manages intellectual property distribution for digitally formatted content (e.g., text, images, audio, video, and animations).

Rights Manager is a working system that encodes license agreement information for intellectual property at a server and distributes the intellectual property to authorized users over the Internet or a campus-wide Intranet along with a Rights Manager-compliant browser. The Rights Manager handles a variety of license agreement types, including public domain, site licensed, controlled simultaneous accesses, and pay-per-use. Rights Manager also manages the functionality available to a client according to the terms of the license agreement; this is accomplished by use of a special browser that enforces the license's terms and which permits or denies client actions such as save, print, display, copy, etc. Access to a particular item of intellectual property, with or without additional functionality, may be made available at no charge, with an overhead charge, or at a royalty plus overhead charge to the client. Rights Manager has been designed to accommodate sufficient flexibility in capturing wide degrees of arbitrariness in charging rules and policies.

The Rights Manager is intended for use by individuals and organizations who function as purveyors of information (publishers, on-line service providers, campus libraries, etc.). The system is capable of managing a wide variety of agreements from an unlimited number of content providers. Rights Manager also permits customization of licensing terms so that individual users or user classes may be defined and given unique access privileges to restricted sets of materials. A relatively common example of this for CWRU would be an agreement to provide (a) view-only capabilities to an electronic journal accessed by an anonymous user

located in the library, (b) display/print/copy access to all on-campus students enrolled in a course for which the digital textbook has been adopted, and (c) full access to faculty for both student- and instructor-versions of digital versions of supplementary textbook materials.

Fundamental to the implementation of Rights Manager are the creation and maintenance of distribution rights, permissions and license agreement databases. These databases express the terms and conditions under which the content purveyor distributes materials to its end-users. Relevant features of Rights Manager include:

- a high degree of granularity for publisher-defined content
- central or distributed management of rights, permissions and licensing databases
- multiple agreement types (e.g., site licensing, limited site licensing and pay-per-use)
- content packaging where rights and permission data are combined with digital format content elements for managed presentation by Web browser "plug-in" modules or helper applications.

Rights Manager maintains a comprehensive set of distribution rights, permissions, and charging information. The premise of Rights Manager is that each publication may be viewed as a compound document. A publication under this definition consists of one or more content elements and media types; each element may be individually managed, as may be required, for instance, in an anthology.

Individual content elements may be defined as broadly or narrowly as required (i.e., the granularity of the elements is defined by the publisher); however, for overall efficiency, each content element should represent a significant and measurable unit of material. Figures, tables, illustrations, and text sections may reasonably be defined as content elements.

To manage the distribution of complete publications or individual content elements, two additional licensing metaphors are implemented. The first of these, a Collection Agreement, is used to specify an agreement between a purveyor and its supplier (e.g., a primary or secondary publisher); this agreement takes the form of a list of publications distributed by the purveyor and the terms and conditions under which these publications may be issued to end-users (one or more Collection Agreements may be defined and simultaneously managed between the purveyor and a customer).

The second abstraction, a Master Agreement, is used to broadly define the rules and conditions that apply to all Collection Agreements between the purveyor and its content supplier. Only one Master Agreement may be defined between the supplier and the institutional customer. In practice, Rights Manager assumes that the purveyor will enter into licensing agreements with its suppliers for the delivery of digitally formatted content. At the time the first license agreement is executed between a supplier and a purveyor, one or more entries are made into the purveyor's Rights Manager databases to define the Master and Collection Agreements. Optionally, Publication and/or Content-Element usage rules may also be defined. Licensed materials may be distributed from the purveyor's site (or perhaps by an authorized service provider); both the content and associated licensing rules are transferred by the supplier to the purveyor for distributed license and content management.

Depending upon the selected delivery option, individual end-users (e.g., faculty members, students or library patrons) may access either a remote server or a local institutional repository to search and request delivery of licensed publications. Depending upon the agreement(s)

between the owner and the purveyor, individual users are assigned access rights and permissions based upon user-IDs, network addresses, or both.

Network or Internet Protocol addresses are used to limit distribution by physical location (e.g., to users accessing the materials from a library, a computer lab or from a local workstation). User identification may be exploited to create limited site-licensing models or individual user agreements (e.g., distributing publications only to students enrolled in Chemistry 432 or, perhaps, to a specific faculty member).

At each of the four permissioning levels (Master Agreement, Collection Agreement, Publication, and Content-Element), access rules and usage privileges may be defined. In general, the access and usage permissions rules are broadly defined at the Master and Collection Agreement level and are refined or restricted at the Publication and Content-Element levels. For example, a general license agreement rule could be defined to specify that by default all licensed text elements may be printed at a some fixed cost, say 10¢ per page; however, high value or core text sections may be individually identified and assessed higher charges, say 20¢ per page, using publication or content element override rules.

When a request for delivery of materials is received, the content rules are evaluated in a bottom-up manner (e.g., content element rules are evaluated before publication rules which are, in turn, evaluated before license agreement rules, etc.). Access and usage privileges are resolved when the system first recognizes a match between the requester's user-ID (or user category) and/or the network address and the permission rules governing the content. Access to the content is only granted when an applicable set of rules specifically granting access permission to the end-user is found; in the case where two or more rules permit access, the rules most favorable to the end-user are selected. Under this approach, site licenses, limited site licenses, individual licensing, and pay-per-use may be simultaneously specified and managed.

The following use of the Rights Manager rules databases is recommended as an initial guideline for Rights Manager implementation:

1) Use Master rules to define the publishing holding company or imprint, the agreement's term (beginning and ending dates), and the general "fair use" guidelines negotiated between a supplier and the purveyor. Because of the current controversy over the definition of "fair use," Rights Manager does not rely upon preprogrammed definitions; rather, the supplier and purveyor may negotiate this definition and create rules as needed. This approach permits "fair use" definitions to be re-defined in response to new standards or regulatory definitions without requiring modifications to Rights Manager itself.

2) Use Collection Agreement rules to define the term (beginning and ending dates) for specific licensing agreements between the supplier and the purveyor. General access and permission rules by user-ID, user category, network address, and media type would be assigned at this level.

3) Use Publication rules to impose any user-ID or user category-specific rules (e.g. permissions for students enrolled in a course for which this publication has been selected as the adopted textbook) or to impose exceptions based on the publication's value.

4) Use Content-Element rules to grant specific end users or user categories access to materials (e.g., define content elements which are supplementary teaching aids for the

instructor) or to impose exceptions based on media type or the value of content elements.

The Rights Manager system does not mandate that licensing agreements exploit user-IDs; however, maximum content protection and flexibility in license agreement specification is achieved when this feature is used. Given that many institutions or consortium customers may not have implemented a robust user authentication system, alternative approaches to uniquely identifying individual users must be considered. While there are a variety of ways in which to address this issue, it is suggested that PIN numbers, assigned by the supplier and distributed by trusted institutional agents at the purveyor's site (e.g., instructors, librarians, bookstore employees or departmental assistants) or embedded within the content be used as the basis for establishing user-IDs and passwords. Using this approach, valid users may enter into registration dialogs to automatically assign user-IDs and passwords in response to a valid PIN "challenge."

While Rights Manager is designed to address all types of multimedia rights, permissions and licensing issues, the current implementation has focused on distribution of traditional print publication media (text and images). Extensions to Rights Manager will be required to address the distribution of full multimedia.

Appendix B: Consortial Standards

MARC

- Enumeration and chronology standards from the serials holding standards of the 853 and 863 fields of MARC
  + Specifies up to 6 levels of enumeration and 4 levels of chronology

  e.g.,

  853 |aVolume|bIssue|i(year)|j(month)
  853 |aVolume|bIssue|cPart|i(year)|j(month)

- Linking from bibliographic records in library catalog via an 856 field
  + URL information appears in subfield "u", anchor text appears in subfield z

  e.g.,

  856 7 |uhttp://beavis.cwru.edu/chemvl|zRetrieve articles from the Chemical Sciences Digital Library

  Would appear as
  Retrieve articles from the Chemical Sciences Digital Library

TIFF

- Most widely used multi-page graphic format
- Support for tagged information ("Copyright", etc.)
- Format is extensible by creating new tags (such as RM rule information, authentication hints, encryption parameters)
- Standard supports multiple kinds of compression

Adobe PDF

- Container for article images
- Page description language
- PDF files are searchable by the Adobe Acrobat browser
- Encryption and security are defined in the standard

SICI (Serial Item and Contribution Identifier)

- SICI Definition (Standards progress, overview, etc.)
- Originally a key part of the indexing structure
- All of the components of the SICI code are stored, so it could be used as a linking mechanism between an article database and the ChemVL Library
- OhioLINK is also very interested in this standard, and is pushing database creators and search engine providers to add SICI number retrieval to citation database and journal article repository systems.
- Future retrieval interfaces into the database: SICI number search form, SICI number search API

  e.g., 0022-2364(199607)121:1<83:TROTCI>2.0.TX;2-I

Appendix C: Equipment Standards for End-Users

Minimum Equipment Required

Hardware: An IBM PC or compatible computer with the following components:

- 80386 processor
- 16MB RAM
- 20MB free disk space
- A video card and display monitor with a resolution of 640 x 480 and 16 colors or shades of gray.

Software:

- Windows® 3.1
- Win32s 1.25
- TCP/IP software suite including a version of Winsock
- Netscape Navigator®2.02
- Adobe Acrobat Exchange® 2.1

Win32s is a software package for Windows 3.1 which is distributed without charge and is available from Microsoft.

The requirement for Adobe Acrobat Exchange, a commercial product which is not distributed without charge, is expected to be relaxed in favor of a requirement for Adobe Acrobat® Reader, a commercial product which is distributed without charge.

The software will also run on newer versions of compatible hardware and/or software.

Recommended Configuration of Equipment

This configuration is recommended for users who will be using the system extensively.

Hardware: A computer with the following components

- Intel Pentium® processor
- 32MB RAM
- 50MB free disk space
- A video card and display monitor with a resolution of 1280 x 1024 and 256 colors or shades of gray.

Software

- Windows NT® 4.0 Workstation
- TCP/IP suite which has been configured for a network connection
- (included in Windows NT)
- Netscape Navigator® 2.02
- Adobe Acrobat Exchange® 2.1

The requirement for Adobe Acrobat Exchange®, a commercial product which is not distributed without charge, is expected to be relaxed in favor of a requirement for Adobe Acrobat® Reader, a commercial product, which is distributed without charge.

Other software options the system has been tested on include:

- IBM OS/2 3.0 Warp Connect® with Win-OS/2
- IBM TCP/IP for Windows 3.1, version 2.1.1
- Windows NT 3.51

Appendix D: Scanning and Workflow

Article Scanning, PDF Conversion and Image Quality Control

The goal of the scan-and-store portion of the project is to develop a complete and tested system of hardware, software and procedures that can be adopted by other members of the consortium with a reasonable investment in equipment, training and personnel. If a system is beyond a consortium member's financial means, it will not be adopted. If a system cannot perform as required, it is a waste of resources.

Our original proposal stressed that all existing scholarly resources, particularly research tools, would remain available to scholars throughout this project. To that end, the scan-and-store process is designed to leave the consortium's existing journal collection intact and accessible.

Scan-and-Store Process Resources

- Scanning workstation, including a computer with sufficient processing and storage capacity, a scanner, and a network connection. Optionally, a second workstation can be used by the scanning supervisor to process the scanned images. The workstation used in this phase of the project includes:

  + Minolta PS-3000 Digital Planetary Scanner
  + Two computers with Pentium 200MHz CPU, 64Mb RAM, 4Gb HD, 21" monitor

  + Windows 3.11 OS (required by other software)
  + Minolta Epic 3000 scanner software
  + Adobe Acrobat Capture, Exchange, and Distiller software
  + Image Alchemy software
  + Network interface cards and TCP/IP software for campus network access

- Scanner operator(s), typically student assistants, with training roughly equivalent to that required for Inter-Library Loan photocopying. Approximately 8 hours of operator labor will be required to process the average 800 pages per day capacity of a single scanning workstation.

- Scanning supervisor, typically a librarian or full-time staff, with training in image quality control, indexing and cataloging, and operation of image processing software. Approximately 3 hours of supervisor labor will be required to process 800 scanned pages per day.

Scan-and-Store Process: Scanner Operator

- Retrieve scan request from system
- Retrieve materials from shelves (enough for two hours of scanning)
- Scan materials and enter basic data into system

  + evaluate size of pages
  + evaluate grayscale/black and white scan mode
  + align material
  + test scan and adjust settings and alignment as necessary
  + scan article
  + log changes and additions to author, title, journal, issue and item data on request form
  + repeat for remaining requested articles

- Transfer scanned image files to Acrobat conversion workstation
- Retrieve next batch of scan requests from system
- Reshelve scanned materials and retrieve next batch of materials

Scan-and-Store Process: Acrobat conversion workstation

- Run Adobe Acrobat Capture to automatically convert sequential scanned image files from single-page TIFF to multi-page Acrobat PDF documents, as they are received from scanner operator
- Retain original TIFF files

Scan-and-Store Process: Scanning Supervisor

- Retrieve request forms for scanned materials
- Open converted PDF files
- Evaluate image quality of converted PDF files

  + scanned article matches request form citation
  + completeness, no clipped margins
  + legibility, especially footnotes and references
  + minimal skewing
  + clarity of grayscale or halftone images
  + appropriate margins, no excessive white space

- Crop fingertips, margin lines, etc., missed by Epic 3000 scanner software

  + retrieve TIFF image file

+ mask unwanted areas
+ re-save TIFF image file
+ repeat PDF conversion
+ evaluate image quality of revised PDF file

- Return unacceptable scans to scanner operator for re-scan or correction
- Evaluate, correct and expand entries in request forms
- Forward corrected PDF files to the database
- Delete TIFF image files from conversion workstation

Notification to and Viewing by User of Availability of Scanned Article

Insertion of the article into the database

- The scanning technician types in the scan request number into a web form.
- The system returns a web form with most of the fields filled in. The technician has an opportunity to correct information from the paging slip before inserting the article into the database.
- The web form contains a "file upload" button that when selected allows the technician to browse the local hard drive for the article PDF file. This file is automatically uploaded to the server when the form is submitted.
- The system inserts the table of contents information into the database and the PDF file to the RightsManager system.

Notification/delivery of article to requester

- E-mail to requester with URL of requested article (in first release)
- No notification (in first release)
- FAX to requester an announcement page with the article URL (proposed future enhancement)
- FAX to requester a copy of the article (proposed future enhancement)

Appendix E: Technical Justification for A Digitization Standard for the Consortium

It is a major premise in the technical underpinnings of the new consortial model that a relatively inexpensive scanner can be located in the major academic libraries of consortium members. After evaluating virtually every scanning device in the market, including some in laboratories

under development, we concluded that the 400 dot-per-inch (dpi) scanner from Minolta was fully adequate for the purpose of scanning all the hundreds of chemical sciences journals in which we were interested. Thus, for our consortium, the Minolta 400 dpi scanner was taken to be the digitization standard. The standard which was adopted preserves 100% of the informational content required by our end-users.

More formally, the standard for digitization in the consortium is defined as follows:

The scanner captures 256 levels of gray in a single-pass with a density of 400 dots-per-inch and converts the gray-scale image to black-and-white using threshold and edge-detection algorithms.

We arrived at this standard by considering our fundamental requirements:

- Handle the smallest significant information presented in the source documents of the chemical sciences literature, which is the lower-case e in super- or sub-scripts as occur in footnotes
- Satisfy both legibility and fidelity to the source document
- Minimize scanning artifacts or "noise" from background
- Operate in the range of preservation scanning
- Be affordable by academic and research libraries

The scanning standard adopted by this project was subjected to tests of footnoted information, and 100% of the occurrences of these characters were captured in both image and character modes and recognized for displaying and searching.

At 400 dpi, the Minolta scanner works in the range of preservation quality scanning as defined by researchers at the Library of Congress (Fleischhauer and Erway, 1992).

We were also cautioned about the problems unique to very high resolution scanning where the scanner produces artifacts or "noise" from imperfections in the paper used. It is a happy note that this was not a problem which we have encountered in this project because the paper used by publishers of chemical sciences journals is coated.

When more is less: Images scanned at 600 dpi require larger file sizes than those scanned at 400 dpi. Thus, 600 dpi is less efficient than 400 dpi. Further, in one series of tests which we conducted, a 600 dpi scanner actually produced an image of effectively lower resolution than 400 dpi. It appears that this loss of information occurs when the scanned image is viewed on a computer screen where there is relatively heavy use of anti-aliasing in the display. When viewed with software which permitted zooming-in for looking at details of the scanned image (which is supported by both PDF and TIFF viewers), the 600 dpi anti-aliased image actually had lower resolution than an image produced from the same source document by the 400 dpi Minolta scanner according to our consortium's digitization standard. With the 600 dpi scanner, the only way for the end-user to see the full resolution was to download the image and then print it out. When a comparison was made of the "soft copy" displayed images, the presentation image quality of 600 dpi was unacceptable to our end-users; the 400 dpi image was just right. Thus, our delivery approach is more useful to the scholar who needs to examine fine details on-screen. We conducted some tests by reconstructing the journal page from the scanned image by printing it out on a Xerox DocuTech 6135 (600 dpi). We found that the smallest fonts actually used and fine details of the articles were uniformly excellent. Interestingly, in many of the tests we

performed, our faculty colleagues judged the end result by their own "acid test:" how good was the scanned image when printed out in comparison with that produced by a photocopier. For the consortium standard, they were satisfied with the result and pleased with the improvement in quality that the 400 dpi scanner provided in comparison with conventional photocopying of the journal page.

. . . . . .

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997